



Source apportionment of exposures to volatile organic compounds: II. Application of receptor models to TEAM study data

Melissa J. Anderson, Eileen P. Daly, Shelly L. Miller*, Jana B. Milford

Department of Mechanical Engineering, University of Colorado at Boulder, UCB 427, Boulder, CO 80309-0427, USA

Received 2 October 2001; accepted 21 March 2002

Abstract

Four receptor-oriented source apportionment models were applied to personal exposure measurements for toxic volatile organic compounds (VOCs). The measurements are from the total exposure assessment methodology studies conducted from 1980 to 1984 in New Jersey (NJ) and California (CA) and the 1987–1990 CA Indoor Exposure study. The receptor models applied are the Chemical Mass Balance model, Principal Component Analysis/Absolute Principal Component Scores, Positive Matrix Factorization, and Graphical Ratio Analysis for Composition Estimates/Source Apportionment by Factors with Explicit Restriction. Major sources of personal exposure to toxic VOCs appear to have been aromatic sources resembling automobile exhaust, gasoline vapor, or environmental tobacco smoke, and a 1,1,1-trichloroethane-dominated source that may be associated with solvent or pesticide use. Drycleaning chemicals, deodorizers or mothballs, and building materials or carpet emissions also appear to have been significant sources of exposure. Source apportionment results from the four models agreed reasonably well for the NJ data. The performance of the models was generally poorer for the CA data, and the corresponding source apportionment results were less consistent across the models. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Hazardous air pollutants; Air toxics; Source attribution

1. Introduction

The majority of toxic volatile organic compounds (VOCs) in ambient air originate from sources that emit to the outdoors, such as motor vehicles (EPA, 2000). It is unclear, however, whether these outdoor sources are the predominant contributors to human exposure to toxic VOCs. Instead, personal activities and indoor sources may be the dominant sources of exposure for many compounds (Wallace, 1991).

Due to the variety and variability of sources involved, it may be more practical to estimate source contributions to toxic VOC exposures using receptor models,

rather than emission-based models (Miller et al., 2001). Receptor modeling has been widely used for estimating source contributions for outdoor air pollution. To our knowledge, however, few studies have applied receptor modeling to personal exposure data (Yakovleva et al., 1999; Anderson et al., 2001).

The primary goals of this study were to identify sources contributing to personal exposure concentrations of toxic VOCs, and estimate the relative contribution of each source to total concentrations. Personal exposure data from the US Environmental Protection Agency's (EPA) TEAM study and the California Air Resources Board's (CARB) California (CA) Indoor Exposure study were analyzed. Four different receptor models were applied to the data: Chemical Mass Balance (CMB) version 7.0 (Watson et al., 1990a, b), Principal Component Analysis/Absolute Principal Component

*Corresponding author. Tel.: +1-303-492-0587; fax: +1-303-492-3498.

E-mail address: shelly.miller@colorado.edu (S.L. Miller).

Scores (PCA/APCS; Thurston and Spengler, 1985), Positive Matrix Factorization (PMF; Paatero and Tapper, 1994) and Graphical Ratio Analysis for Composition Estimates/Source Apportionment by Factors with Explicit Restriction (GRACE/SAFER; Henry et al., 1994), as incorporated in the UNMIX model (version 2; Henry, 2000). In Part I of this paper, Miller et al. (2002) applied these four models to a simulated data set of VOC exposures. Anderson et al. (2001) previously analyzed the TEAM and CARB data using PMF. This study examines whether results similar to those obtained with PMF are found using alternative receptor models.

2. Data

Personal exposure data for residents in Elizabeth and Bayonne, New Jersey (NJ) were accessed from the Total Human Exposure Risk database and Advanced Simulation Environment, version 1.2 (Pandian et al., 1989; NERL, 2000). This database includes data from the main TEAM studies, which measured 24-h exposures of 600 people in NJ, CA, North Dakota and North Carolina (NC) to various toxic chemicals in air and drinking water. Data for residents of Los Angeles, Pittsburgh, and Antioch, CA were accessed from the Californian Exposures Database (CED) (Clayton and Perritt, 1993). The CARB studies were carried out using the same general procedures as the TEAM studies.

Exposures were measured using personal monitors, which collected two 12-h samples, representing over-night and daytime exposures. More information on these studies is provided elsewhere (Pandian et al., 1989; Clayton and Perritt, 1993; Wallace, 1987). A summary of the measured personal exposure concentrations is presented in Table 1. For this analysis, data were combined across study locations within each state and across seasons. The influence on source apportionment results of geographic and seasonal variations in the data was examined by Anderson et al. (2001). PMF and CMB results for outdoor data from the TEAM and CARB studies are presented in Anderson et al. (2001) and Anderson (2001).

Compounds and participants were included in this analysis only if more than 60% of the corresponding concentrations were reported. Missing values were filled in using the median value for the compound across all participants. Table 1 shows that relatively few data points were missing after compounds and participants not meeting the 60% completeness criterion were excluded. To decrease the effect of outliers on the results, only participants with total exposure concentrations below $2000 \mu\text{g}/\text{m}^3$ were included in the NJ data set, excluding 4% of the participants. This cutoff concentration was an obvious divider between the most extreme concentrations and the remaining values. In the CA data, only one participant had a total exposure concentration above $2000 \mu\text{g}/\text{m}^3$, so all participants were included in the CA analysis.

Table 1
Integrated 24-h personal exposure concentrations ($\mu\text{g}/\text{m}^3$) of compounds included in this study

Compound (abbreviation)	NJ data		CA data	
	No. Obs.	GM (GSD)	No. Obs.	GM (GSD)
Benzene (BNZ) ^a	506	12 (4.1)	296	9.7 (2.1)
Carbon tetrachloride (TET) ^a	505	1.4 (2.6)	299	0.73 (1.7)
Chlorobenzene (CBZ) ^a	507	0.33 (2.2)	287	0.063 (2.2)
Chloroform (CFM) ^a	506	2.9 (3.7)	298	0.70 (4.1)
1,2-Dichlorobenzene (ODB)	507	0.35 (2.8)	—	—
1,4-Dichlorobenzene (PDB) ^a	506	5.4 (5.7)	—	—
1,2-Dichloroethane (DCA)	507	0.54 (2.7)	283	0.15 (2.5)
Ethylbenzene (EBZ) ^a	506	7.4 (2.9)	304	5.1 (2.6)
<i>n</i> -Octane (OCT)	—	—	302	3.3 (2.3)
Styrene (STR) ^a	507	2.0 (2.7)	301	1.6 (2.9)
Tetrachloroethylene (PRC) ^a	508	7.8 (3.2)	291	4.9 (3.3)
1,1,1-Trichloroethane (TCA)	501	16 (3.9)	305	18 (3.4)
Trichloroethylene (TCE) ^a	508	3.1 (3.4)	298	1.1 (6.2)
<i>o</i> -Xylene (OXY) ^a	506	6.8 (2.6)	305	6.5 (2.7)
<i>m,p</i> -Xylene (MPX) ^a	507	19 (2.8)	305	17 (2.3)
<i>n</i> -Decane (DEC)	—	—	295	2.2 (3.2)
<i>n</i> -Dodecane (DOD)	—	—	288	1.3 (3.1)
<i>n</i> -Undecane (UND)	—	—	295	2.3 (3.1)
α -Pinene (PIN)	—	—	301	2.2 (3.0)

^a On current federal HAPs list.

3. Models

PCA/APCS, PMF, and UNMIX were applied to the NJ and CA data to estimate source profiles and contributions. CMB was used to estimate source contributions, using profiles that were input to the model. Miller et al. (2002) provide a brief description of each model. Details that are specific to this application are discussed below.

CMB uses the chemical composition of receptor concentrations to estimate the contributions of different sources using weighted least-squares linear regression (Watson et al., 1990a,b). The model input consists of chemical composition profiles for likely sources. The user determines the best number and combination of source profiles to fit the data based on goodness of fit measures, including R^2 and χ^2 values, and the ratio of calculated to actual total concentrations (% mass). Different source profiles can be selected to optimally fit the exposure data for each individual participant. Here, however, a single set of profiles was selected that best fit the data overall. A total of 28 relevant source profiles and associated error estimates were compiled and tested in the CMB model (Anderson, 2001). The subsets of these profiles that gave the best fits to the CA and NJ data are shown in the results section. Although source profiles for the locations and dates of the TEAM studies were not available, the profiles used generally represent emissions in the 1980s and early 1990s. Some of the profiles compiled were measured at the point of exposure and some were emission profiles. Ideally, exposure profiles that take aging and chemical degradation into account would be available for all sources. Anderson (2001) documents the full set of profiles tested with the CMB model.

In PCA, the matrix of pairwise correlations among compound concentrations is decomposed into eigenvectors, which are then sorted by descending order of the corresponding eigenvalues. In this study, eigenvectors were carried through the analysis if the associated eigenvalues were close to or > 1 . Varimax rotation was applied to these eigenvectors. Because factor scores given by PCA are correlated with but not proportional to source contributions, quantitative estimates of source contributions and profiles associated with each factor are derived by calculating APCS and then regressing the total VOC concentrations on these scores (Thurston and Spengler, 1985). PCA is based on the correlation structure of the observations, and so cannot reliably handle missing data. In this study, however, PCA was applied to the same data set as the other methods, because only a small number of missing observations had to be filled in.

PMF, another multivariate method, solves a constrained, weighted least-squares optimization problem to factor a data set into profile and factor score matrices.

The number of factors in the solution is chosen by the user. Error estimates for each observation are used in the weighted least-squares solution to downweight less reliable data. The error model described by Miller et al. (2002) was used in this study, except that for observations that were filled in, the error was set to two times the median value of the compound. The number of factors in the PMF model was chosen based on the normalized sum of squares of the errors in individual species concentrations (Q -value), the distribution of normalized residuals for individual compounds, and the results of a linear regression analysis of the factor scores versus the sum of the measured VOC concentrations, in which we sought to avoid negative regression coefficients and to optimize the R^2 and % mass metrics for the regression (Anderson et al., 2001). The recommendations of Paatero and Tapper (1994) were followed for the rotational freedom (FPEAK = 0), outlier threshold distance ($\alpha = 4.0$) and choice of the robust solution mode.

The GRACE/SAFER method, incorporated in the UNMIX model, obtains source compositions by a combination of graphical analysis and multivariate receptor modeling (Henry and Kim, 1990). The model includes physical constraints, such as non-negativity of source compositions and contributions, as an integral part of the factor profile derivation procedure. The user specifies a tracer, which is a compound contributed predominantly by one source, and the number of factors. Solutions were attempted with subsets of compounds, different tracers and different numbers of factors until a feasible solution was found that optimized model performance. The optimal number of factors was chosen based on the minimum R^2 value for any single compound (Min R^2) and the minimum signal-to-noise ratio for any compound (Min S/N) (Henry, 2000). Ultimately, eight compounds were included in the solution for NJ and nine compounds in the solution for CA, with tetrachloroethylene designated as the tracer in both cases. The filter parameters that determine how stringently the non-negativity constraint is applied and the parameters that downweight extreme values were set to the same values used by Miller et al. (2002).

Applying UNMIX to the full data sets for NJ and CA did not produce meaningful factors, due to outliers. To identify possible outliers, concentrations of each compound were plotted against concentrations of TCA, one of the most dominant and variable compounds in the data sets. In this way, one outlier was identified and removed from the CA data set. Due to the variability of the NJ data set, locating outliers using this approach was difficult. Therefore the NJ data were trimmed to exclude participants if the natural logarithm of their total exposure concentration was more than $\pm 2\sigma$ (standard deviations) of the mean natural logarithm.

About 5% of the participants were thus excluded, with UNMIX applied to the data for the 482 remaining participants.

4. Results

Source apportionment was performed for the NJ and CA personal exposure data sets using CMB, PCA/APCS, PMF, and UNMIX. Factor profiles from each model are presented in Figs. 1–4 for NJ and Figs. 5–8 for CA. The profiles shown for the CMB model are input to the analysis, so the CMB figures also include profile names and references. The height of the bars in these figures represents the compound's mass fraction in

the factor. The error bars shown for the PMF and UNMIX factors represent $\pm 1\sigma$ errors in the profiles, as output by each model. Error bars on the CMB results represent $\pm 1\sigma$ of variability in the measured input profiles, as reported in the references. Error estimates for APCS factor profiles were not calculated. Summary performance measures provided by each model are given in the figure captions. Anderson (2001) presents additional performance measures for the PMF and CMB models.

With CMB (Fig. 1), six source profiles were judged to best fit the NJ personal exposure data, resulting in an average R^2 value of 0.82, across all participants. Performance measures for individual compounds given in Anderson (2001) indicate that the CMB model

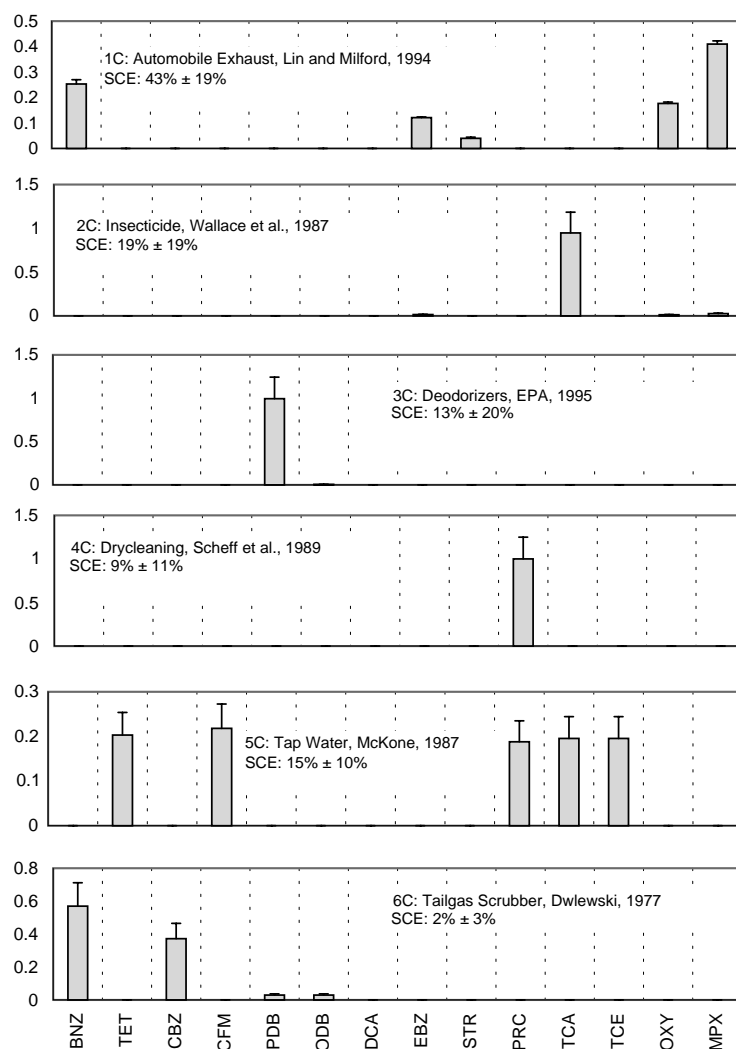


Fig. 1. CMB input source profiles (mass fractions) that best fit the NJ personal exposures ($n = 508$) with average and standard deviation of source contribution estimates (SCE) across all participants. Average and standard deviation of performance statistics across participants: $R^2 = 0.82 \pm 0.13$; $\chi^2 = 4.0 \pm 3.6$; % mass = $90\% \pm 36\%$.

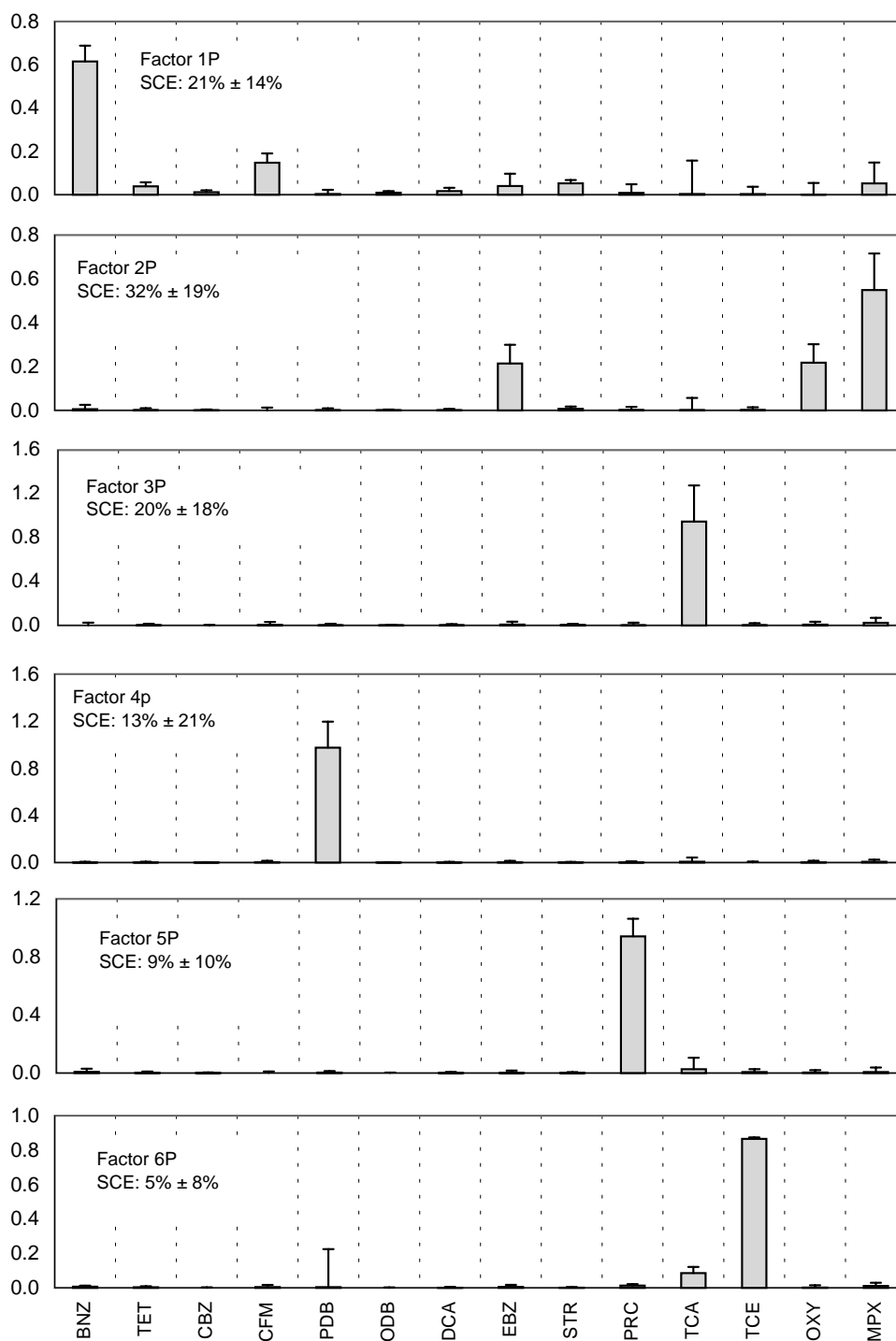


Fig. 2. PMF factorization results (mass fractions) for NJ personal exposures ($n = 508$). Performance statistics: $R^2 = 0.97$ and % mass = $105\% \pm 12\%$ for total VOC concentrations; $Q = 16305$ for individual species concentrations.

overpredicts BNZ and TET by more than a factor of two, on average. DCA is not included in any of the source profiles used with CMB; hence, it is not fit at all.

Because measured DCA concentrations are low, this omission does not significantly affect the model's overall performance. On average, modeled concentrations of

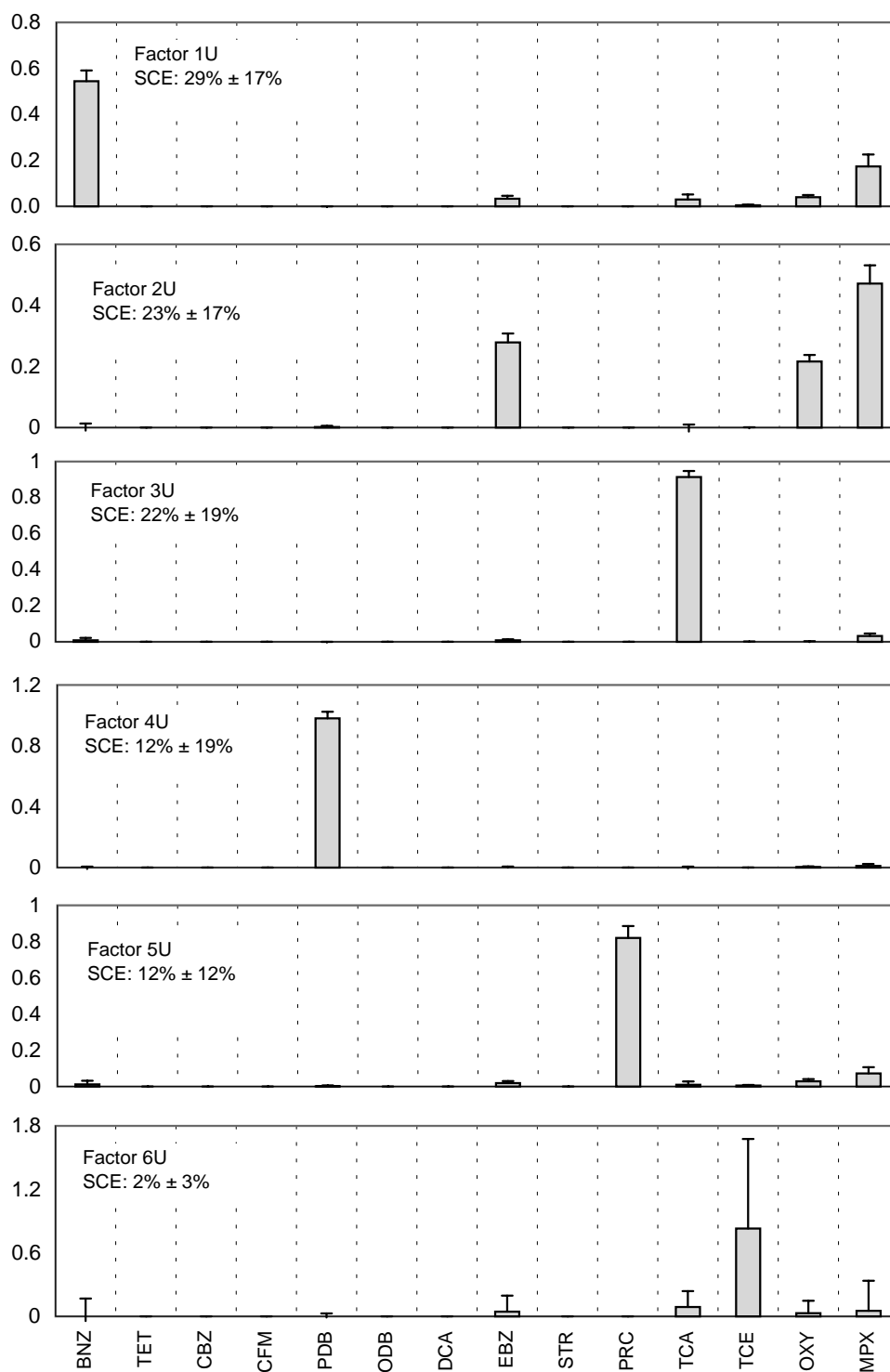


Fig. 3. UNMIX factorization results (mass fractions) for NJ personal exposures ($n = 482$). TET, CBZ, CFM, ODB, DCA and STR were omitted from the UNMIX model; spaces are left for them to aid comparison with the other models. Performance statistics: Min $R^2 = 0.88$ for any species; Min S/N = 2.46 for any species; % mass = $99\% \pm 13\%$.

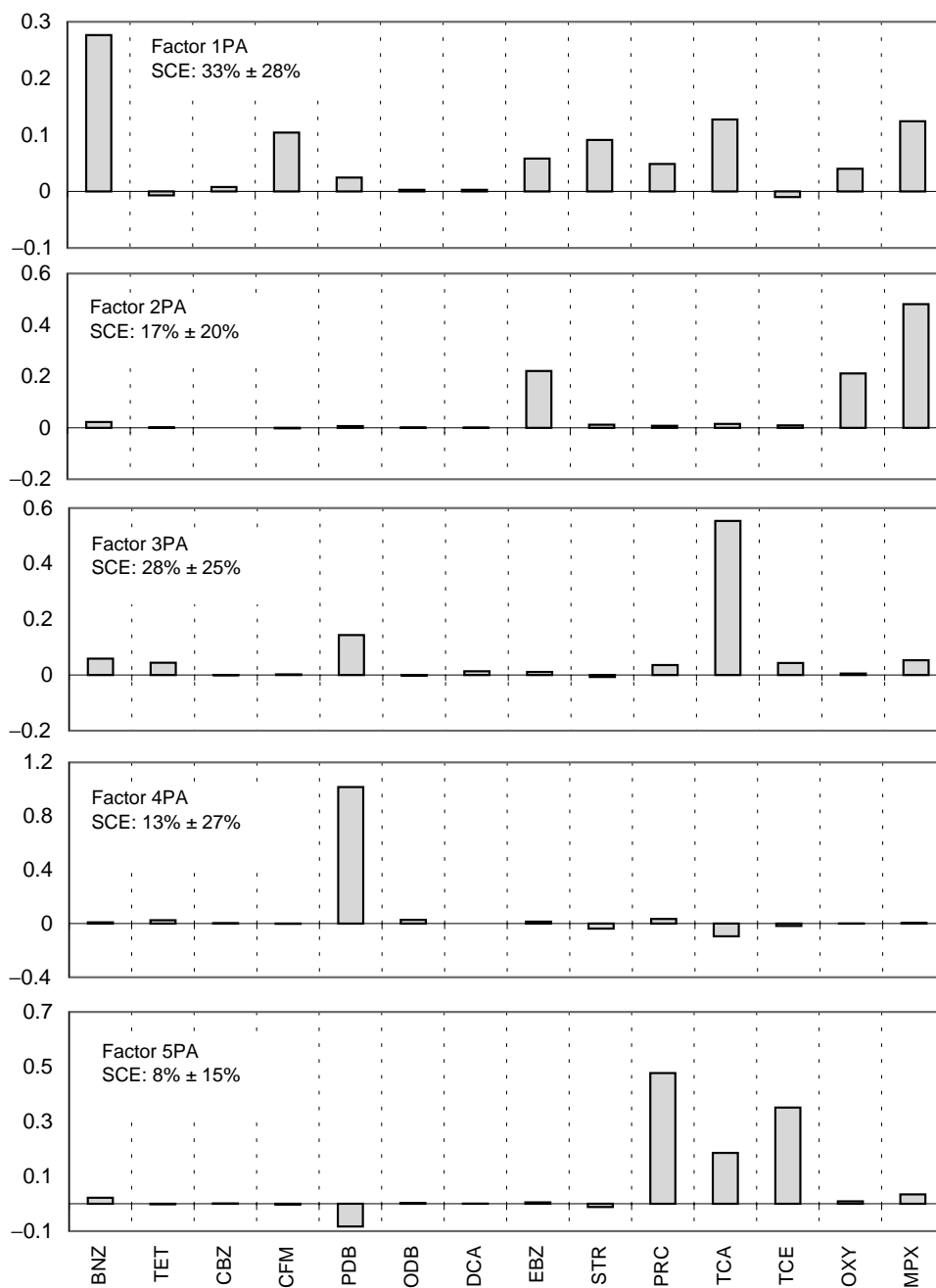


Fig. 4. PCA/APCS factorization results (mass fractions) for NJ personal exposures ($n = 508$). Performance statistics: $R^2 = 0.87$ for total VOC concentrations.

the other compounds are within a factor of two of the measured concentrations.

Six factors were chosen as the optimal number for both the PMF and UNMIX models (Figs. 2 and 3, respectively). Eight compounds: BNZ, PDB, EBZ, PRC, TCA, TCE, OXY and MPX were included in the UNMIX solution. Detailed results for PMF (Anderson,

2001) indicate that more than 80% of the residuals were within the $\pm 2\sigma$ uncertainty bounds on the observed values for all of the compounds except for CFM and STR, for which 69% and 63% of the residuals, respectively, fell within these limits. With UNMIX, the minimum R^2 value of 0.88 and minimum signal-to-noise ratio of 2.5 are both above the respective thresholds of

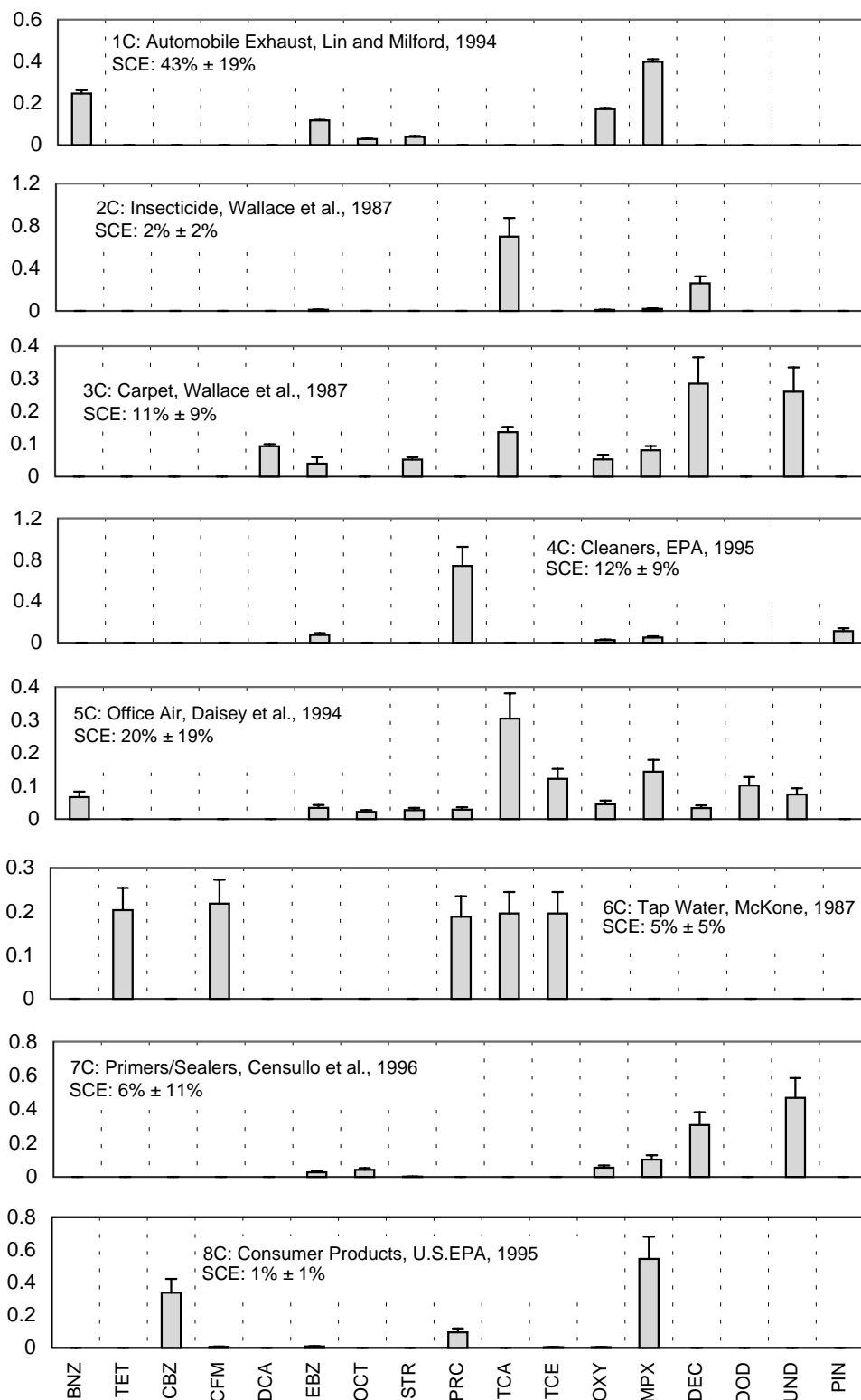


Fig. 5. CMB input source profiles (mass fractions) that best fit the CA personal exposures ($n = 305$). Average and standard deviation of performance statistics across participants: $R^2 = 0.81 \pm 0.09$; $\chi^2 = 12 \pm 6.3$; % mass = $82\% \pm 18\%$.

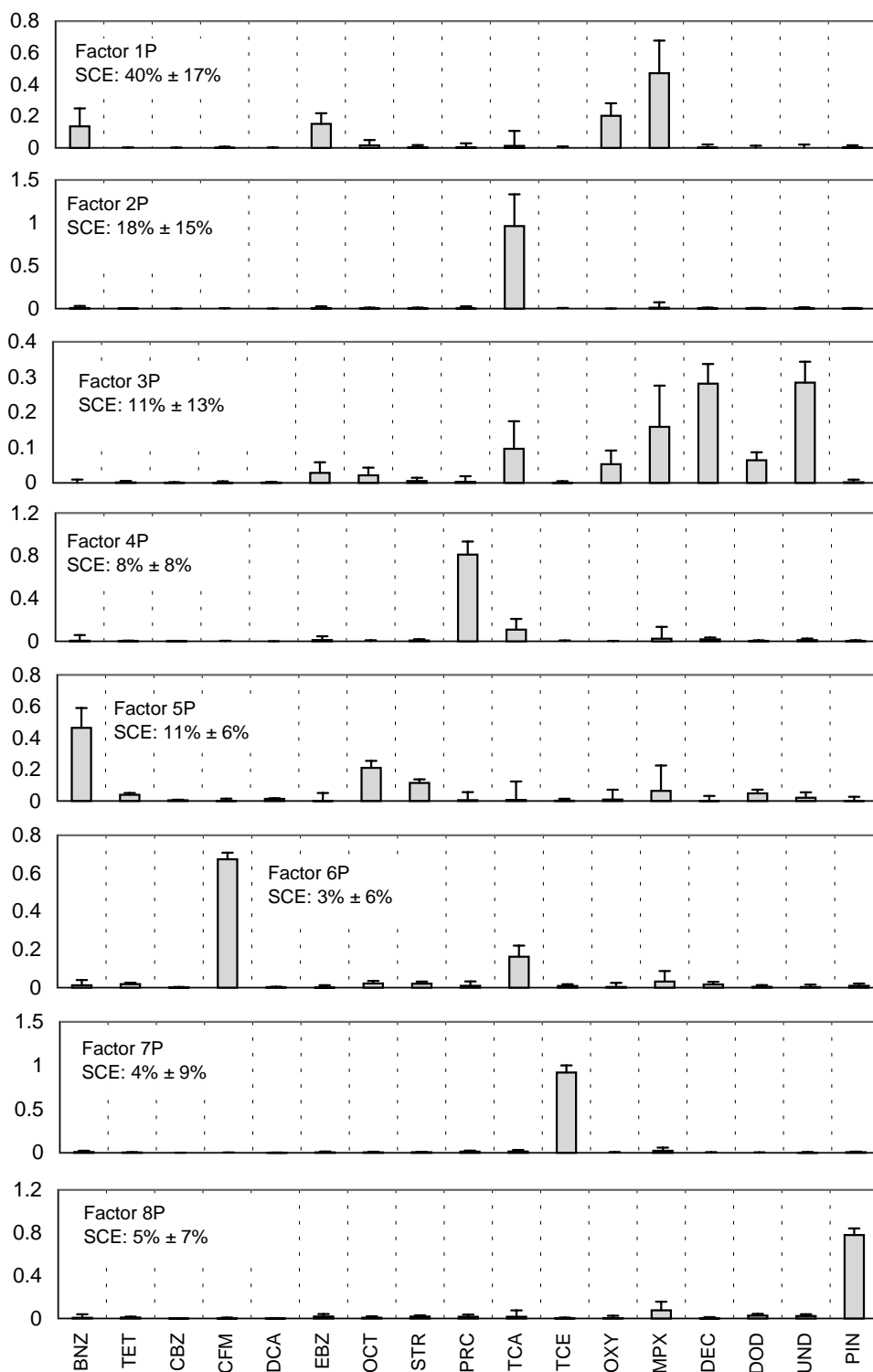


Fig. 6. PMF factorization results (mass fractions) for the CA personal exposures ($n = 305$). Performance statistics: $R^2 = 0.98$ and % mass = $100\% \pm 11\%$ for total VOC concentrations; $Q = 14351$ for individual species concentrations.

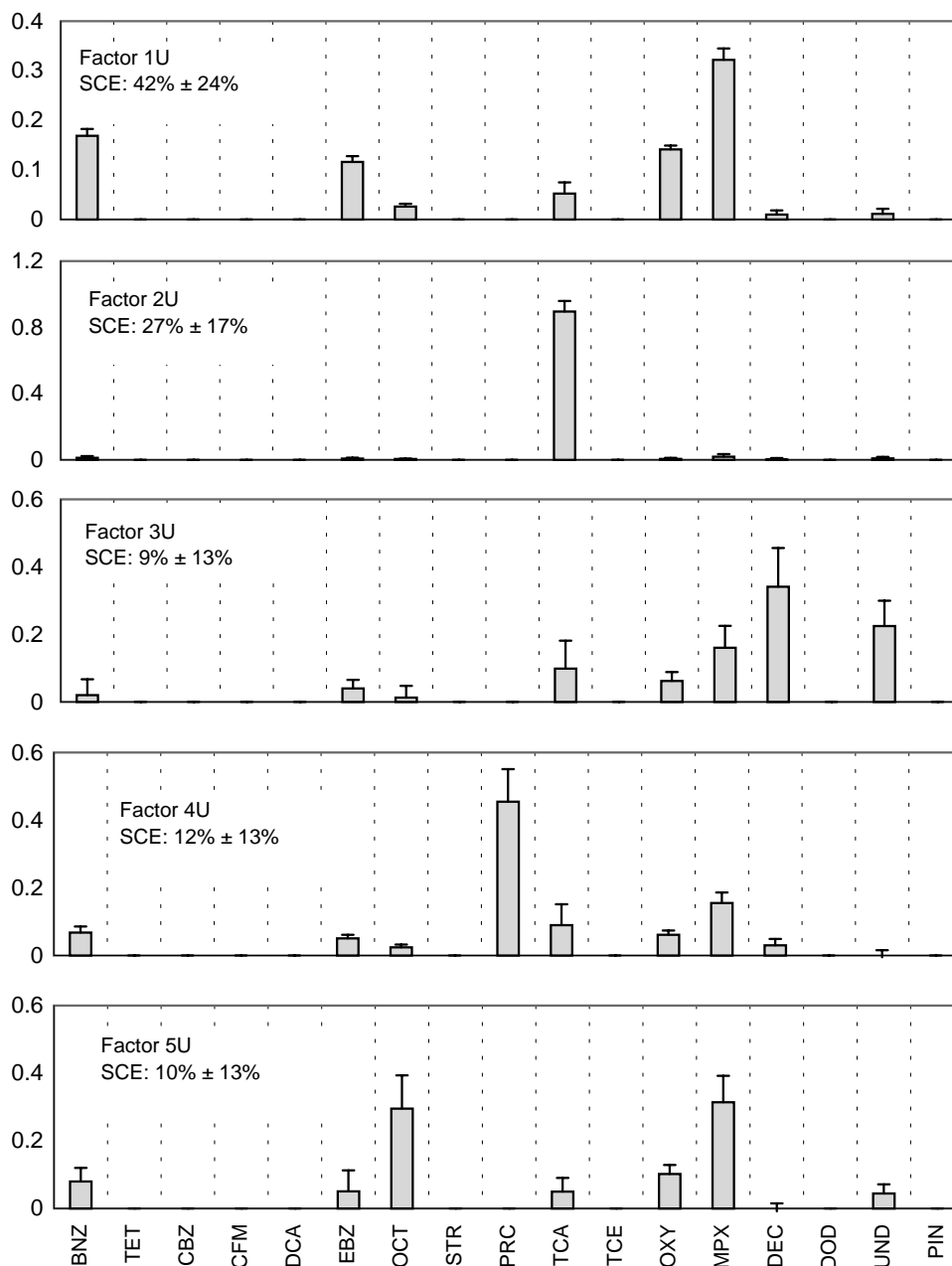


Fig. 7. UNMIX factorization results (mass fractions) for the CA personal data ($n = 304$). TET, CBZ, CFM, DCA, STR, TCE, DOD and PIN were omitted from the UNMIX model; spaces are left for them to aid comparison with the other models. Performance statistics: Min $R^2 = 0.65$ for any species; Min S/N = 1.20 for any species; % mass = $92\% \pm 20\%$.

0.80 and 2.0 recommended for a good model (Henry, 2000).

Five factors were carried through the PCA/APCS analysis of the NJ data (Fig. 4), accounting for 60% of the variance in the exposure concentration data for individual compounds. Inclusion of additional factors resulted in unacceptably large negative values in the

factor profiles determined from APCS. The R^2 value for the regression of total VOC concentrations to the absolute principal component scores was 0.87.

Table 2 compares source contribution estimates for profiles estimated or used by the four models for NJ. The average and standard deviation of source contributions across participants is shown. Results for similar

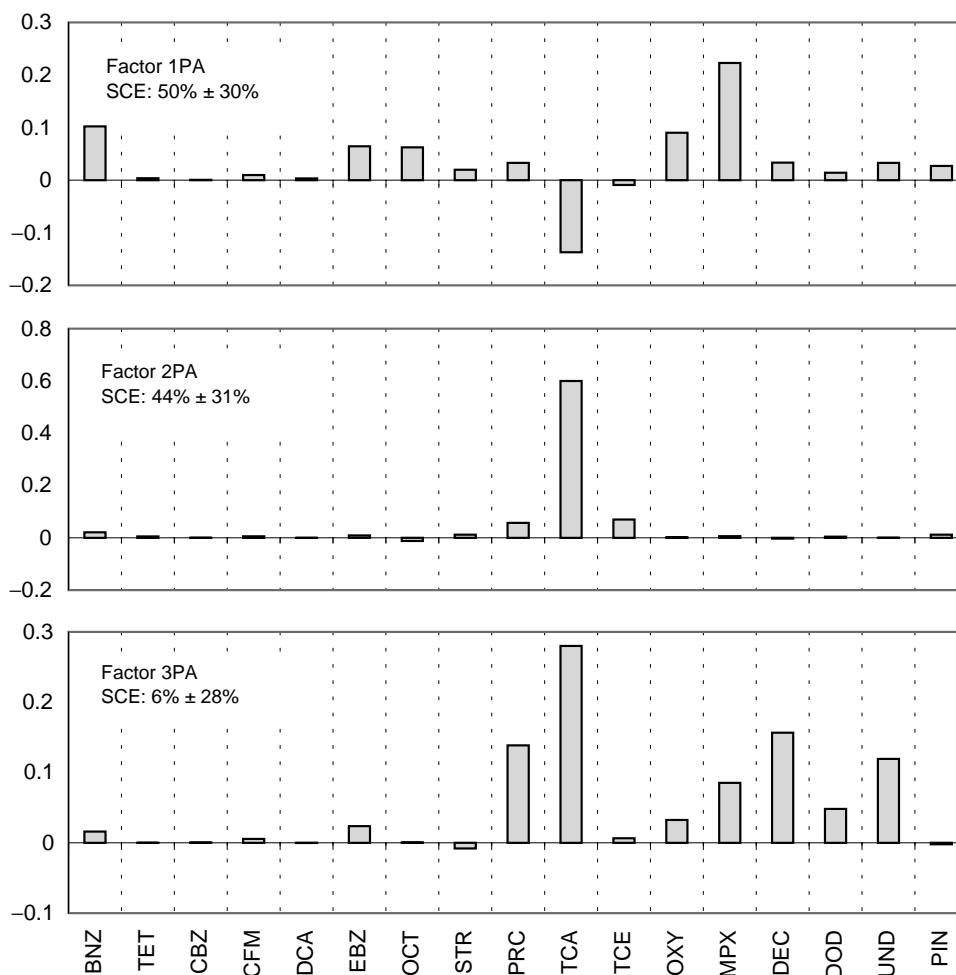


Fig. 8. PCA/APCS factorization results (mass fractions) for CA personal exposures ($n = 305$). Performance statistics: $R^2 = 0.82$ for total VOC concentrations.

profiles are lined up next to each other, with major compounds present in the profiles indicated in the first column. A blank entry indicates that no similar profile was found for that model. Note that UNMIX results are not expected to exactly match those from the other models because it was applied to a trimmed data set.

For the NJ exposures, the largest source contributions are from aromatics-dominated factors. The PMF, UNMIX and PCA/APCS models all extract a factor dominated by BNZ and a second factor with EBZ, OXY and MPX. The first factor contributes an average of 21–33% of the total VOC concentrations, depending on the model, while the second factor contributes from 17% to 32%. The CMB model uses an automobile exhaust profile that resembles a combination of the first and second factors from the other models and contributes an average of 43% to total VOC concentrations. The next factor in all the model results is dominated by TCA, and

contributes from 19% to 28% of total VOC concentrations. All four models included a PDB factor, which contributes from 12% to 13% to the personal exposure concentrations in NJ. The next two factors found by PMF and UNMIX are dominated by TCE and PRC, and contribute from 2% to 5% and from 9% to 12% to personal exposures, respectively. The CMB model also uses a PRC-dominated profile, which contributes 9%. The final PCA/APCS profile resembles a combination of the TCE and PRC profiles from PMF and UNMIX.

For the CA data, eight source profiles were used in the CMB model, giving an average R^2 value of 0.81 (Fig. 5). Eight factors were chosen for the PMF model (Fig. 6) and five for UNMIX (Fig. 7). Nine compounds: BNZ, EBZ, OCT, PRC, TCA, OXY, MPX, DEC and UND were included in the UNMIX solution. Three factors were carried through the PCA/APCS model (Fig. 8) accounting for 49% of the variance in the data.

Table 2

Average source contribution estimates (SCE, %) for NJ personal exposure concentrations

Major compounds	CMB		PMF		UNMIX		PCA/APCS	
	Profile	SCE	Profile	SCE	Profile	SCE	Profile	SCE
BNZ, EBZ, xylenes	1C	43 ± 14						
BNZ			1P	21 ± 14	1U	29 ± 17	1PA	33 ± 28
EBZ, xylenes			2P	32 ± 19	2U	23 ± 17	2PA	17 ± 20
TCA	2C	19 ± 19	3P	20 ± 18	3U	22 ± 19	3PA	28 ± 25
PDB	3C	13 ± 20	4P	13 ± 21	4U	12 ± 19	4PA	13 ± 27
PRC	4C	9 ± 11	5P	9 ± 10	5U	12 ± 12		
TCE			6P	5 ± 8	6U	2 ± 3		
TCE, PRC							5PA	8 ± 15
TET, CFM, PRC, TCA, TCE	5C	15 ± 10						
BNZ, CBZ	6C	2 ± 3						

Standard deviations representing variability across participants are shown.

Table 3

Average source contribution estimates (SCE, %) for CA personal exposure concentrations

Major compounds	CMB		PMF		UNMIX		PCA/APCS	
	Profile	SCE	Profile	SCE	Profile	SCE	Profile	SCE
BNZ, EBZ, xylenes	1C	43 ± 19	1P	40 ± 17	1U	42 ± 24	1PA	50 ± 30
TCA	2C	2 ± 2	2P	18 ± 15	2U	27 ± 17	2PA	44 ± 31
TCA, MPX, DEC, DOD, UND	3C	11 ± 9	3P	11 ± 13	3U	9 ± 13	3PA ^a	6 ± 28
PRC	4C	12 ± 9	4P	8 ± 8	4U	12 ± 13		
TCA, TCE, OXY, MPX, DOD, UND	5C	20 ± 19						
TET, CFM, PRC, TCA, TCE	6C	5 ± 5						
DEC, UND	7C	6 ± 11						
CBZ, MPX	8C	1 ± 1						
BNZ, OCT, OXY, MPX			5P	11 ± 6				
CFM			6P	3 ± 6				
TCE			7P	4 ± 9				
PIN			8P	5 ± 7				
BNZ, OCT, xylenes					5U	10 ± 13		

Standard deviations representing variability across participants are shown.

^a Also includes PRC as a major compound.

Although eigenvalues corresponding to three other factors were close to 1.0, their inclusion in the APCS calculation led to large negative values in the factor profiles.

Based on the aggregate performance measures provided by each model and shown in the figure captions, the overall performance of the PMF model was similar for the CA and NJ data sets, but the other three models performed worse for CA than for NJ. The high average χ^2 value of 12 for the CA CMB application suggests that concentrations of one or more species are not explained well by the model. The more detailed results presented by Anderson (2001) indicate that OCT is underestimated by a factor of two, on average, and that TCE is overestimated by almost a factor of three. Residuals from PMF for BNZ, TET, DCA, OCT, STR and PRC

were within the $\pm 2\sigma$ uncertainty limits for only about 70% of the observations, with residuals for the other compounds within these limits for at least 80% of the observations (Anderson, 2001). With UNMIX, the minimum R^2 value of 0.65 and minimum signal-to-noise ratio of 1.2 are below the recommended values of 0.80 and 2.0, respectively. We were unable to improve UNMIX's performance for the CA data and still produce meaningful factors. Finally, the R^2 value for the APCS model fit to total VOC exposure concentrations in CA was 0.82, compared to 0.87 for NJ.

Table 3 compares source contribution estimates for the CA data corresponding to the factors shown in Figs. 5–8. The results are presented in the same format as in Table 2. All the four models found or used a factor that was dominated by BNZ, EBZ and xylenes. On

average, this factor contributed from 42% to 50% of the total exposure concentrations in CA. A second factor in all the model results was dominated by TCA, but the corresponding source contributions ranged from 2% for the CMB model to 44% for PCA/APCS. Note that TCA is also prominent in the indoor air profile (5C) used with the CMB model, which contributes 20% to total VOC concentrations. The relatively high PCA/APCS estimate of the average TCA source contribution may have been strongly influenced by a few very high TCA concentrations. The outlier removed from the data for UNMIX, which had a very high TCA concentration, was not removed for the other models. Next, PMF, UNMIX and CMB include a similar factor containing TCA, MPX and alkanes that contribute 9–11% of total VOC concentrations, on average. The third profile obtained with PCA/APCS is similar, but also includes PRC. The average source contribution estimate for this factor is 6%. The final point of agreement between the PMF, UNMIX and CMB models is the inclusion of a separate PRC factor, which contributes from 8% to 12% of total VOC concentrations.

5. Discussion

The factors contributing most to the total personal exposure concentrations in NJ and CA and appearing across multiple models were interpreted as six different sources or combinations of sources, based on a qualitative comparison to emissions or exposure profiles reported in the literature. Precise matches with profiles reported in the literature were not expected due to variability in the reported profiles and the fact that VOCs can undergo chemical degradation from the time they are emitted. The six most significant sources appear to be:

- gasoline vapors (Scheff et al., 1991; Rappaport et al., 1987), automobile exhaust (Scheff et al., 1991; Lin and Milford, 1994), or ETS (Daisey et al., 1994) (aromatics, including BNZ, EBZ, and xylenes);
- solvents and insecticides (TCA) (Wallace et al., 1989);
- drycleaning chemicals (PRC) (Wallace et al., 1989; Wallace, 1989);
- tap water (TCE) (McKone, 1987);
- deodorizers or mothballs (PDB) (Wallace, 1989);
- building materials and carpeting (higher alkanes and xylenes) (Mølhave, 1982; Wallace et al., 1987).

Factors corresponding to the first four of these sources appeared in both data sets. The PDB factor appeared only in the NJ data. This compound was not included in the CA analysis due to an inadequate number of measurements. The factor associated with

building materials and carpeting was only extracted from the CA data. The alkanes were not measured in the NJ study. The factor labeled tap water was less consistent across models than the other major factors. TCE appeared as the dominant compound in one of the factors given by PMF and UNMIX for NJ and by PMF for CA, was present along with TCA and PRC in a PCA/APCS factor for NJ, and was a significant component in the tap water profile included in the CMB model for both CA and NJ.

Each of the four models identified one or two aromatics factors as the primary contributor to both the NJ and CA exposure concentrations. For the NJ data, factors 1PA, 1P, and 1U are dominated by BNZ while 2PA, 2P, and 2U consist mainly of EBZ and xylenes. The sums of these two factors from each of these models are shown in Fig. 9, and are compared with the automobile exhaust profile (1C) used with the CMB model for both NJ and CA. The summed factors extracted from the NJ data are similar to the CMB profile, and in turn are similar to factors 1PA, 1P and 1U extracted from the CA personal exposure data. Although the aromatics-dominated profile used with the CMB model was for automobile exhaust (Scheff et al., 1991), gasoline vapors (Scheff et al., 1991; Rappaport et al., 1987) and ETS (Daisey et al., 1994) also have similar profiles with respect to the compounds available for use in this study (see Miller et al., 2002, Fig. 5). Thus it is likely that the aromatics factors reflect exposure to a combination of aromatics-dominated sources, contributing from 43% to 53% of personal exposure to toxic VOCs in NJ, on average, and from 42% to 50% in CA.

In addition to the aromatics-dominated factors, another multi-component factor extracted from the CA data that was relatively consistent across models appears to be building materials and/or carpeting profile. CA factors 3P and 3U are similar to CMB profile 3C, for emissions from carpeting (Wallace et al., 1987). The dominant compounds on factor 3PA are also similar, although it uniquely includes a significant amount of PRC.

Overall, PMF and UNMIX agree especially well for the NJ data set, with both models extracting six similar factors. Table 2 shows that the average source contribution estimates for these two sets of factors are also similar. For NJ, results from PCA/APCS also compare reasonably well to those from PMF and UNMIX, although PCA combines PRC and TCE, which the other models separate. PCA/APCS also attributes a relatively high average source contribution to the BNZ-dominated factor (1PA) and a correspondingly lower contribution to the second aromatics factor (2PA). The automobile exhaust profile used in the CMB model appears to be separated by the other models into two profiles. Their summed source contribution estimates are about 25%

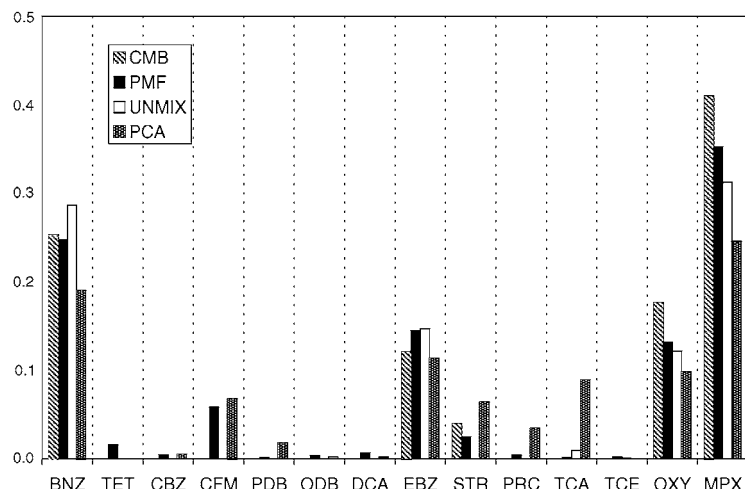


Fig. 9. Comparison of aromatics source profiles (mass fractions) from the four models. The CMB input profile is for automobile exhaust. The sums of the profiles extracted from the NJ data for factors 1PA and 2PA, 1P and 2P, and 1U and 2U are shown for PCA/APCS, PMF and UNMIX, respectively.

higher than the automobile exhaust contribution from CMB.

The models give less consistent results for personal exposures in CA than they do for NJ. Profiles similar to CMB profiles 1C–4C are extracted by PMF and UNMIX. The average source contribution estimates match well for three of the four common factors, but CMB gives a low SCE for the TCA-dominated factor. Although the factor profiles are similar, source contribution estimates for the first two PCA/APCS factors appear high compared to the corresponding factor contributions from the other models. Beyond this, the CA results from the four models diverge sharply.

There are several reasons why the results obtained with the different models may disagree. First, exposure profiles specific to the time period and locations included in the TEAM and CARB studies were not available in the literature, so discrepancies were expected between the input profiles used with CMB and the profiles extracted by the other models. Some of the VOCs included in the study undergo decay in the atmosphere, so the exposure data likely violate the assumption made in the CMB model that source compositions remain constant.

The PCA/APCS model differs fundamentally from the UNMIX and PMF models in not having a non-negativity constraint. As mentioned above, we limited the number of factors carried through the PCA/APCS analysis in order to avoid large negative values in the factor profiles, with the result that less of the variability in the data was explained.

PMF, unlike PCA/APCS or UNMIX, utilizes point-by-point estimates of errors in the data, allowing

downweighting of missing observations that have been filled in, values that are close to the detection limits, and outliers. UNMIX provides a relatively coarse means of downweighting outliers, through two adjustable weighting parameters.

Based on the formulation of the models, and as seen in the results for the simulated data reported by Miller et al. (2002), none of the models are expected to be able to discriminate between sources with similar compositions. Likewise, UNMIX and PCA may be confounded by high correlations between sources of exposure, which could be introduced by common meteorological factors or the presence of the sources in the same microenvironments. The results for the simulated data also suggest that factors extracted by PMF, PCA and UNMIX that are dominated by a single compound may reflect a high degree of variability in the amount of that compound in one or more multi-component sources, rather than a pure source (Miller et al., 2002). This may be the case with BNZ and TCE in the NJ data, for example.

The performance measures reported by the CMB, PMF, PCA/APCS and UNMIX models are useful for comparing results obtained with alternative configurations of a given model (e.g., selecting different input profiles with CMB or different numbers of factors with PMF, PCA/APCS and UNMIX). Because the models report different performance measures, however, they are less useful for cross-model comparisons. In particular, the current version of UNMIX reports only the minimum R^2 and S/N values for the worst-fit compound included in the model, but does not provide any measure of model performance for the other compounds.

6. Conclusions

Of the four receptor models compared here, PMF and UNMIX showed the closest agreement, extracting six similar factors from the NJ personal exposure data and four similar factors from the CA data. Between the two, PMF has the advantage of reporting more complete information on model performance than the current version of UNMIX. The lack of a non-negativity constraint is a significant limitation of both PCA/APCS and CMB. Nevertheless, results from CMB were useful in helping to interpret the factors extracted by the other models.

Across the models, the results suggest that in the late 1980s important sources of personal exposure to toxic VOCs included an aromatics source resembling automobile exhaust, gasoline vapors or ETS; a TCA source such as solvent or insecticide use; a PDB source such as mothballs; PRC from dry cleaning; and in CA only, higher alkanes and xylenes from building materials. Tap water was also tentatively identified as a source of exposure, but the corresponding profiles and source contribution estimates differed more across models than those for the other sources. Source profiles found in the literature for automobile exhaust, gasoline vapors and ETS are too similar with respect to the compounds included in this study to be able to distinguish them. Since the aromatics-dominated source appears so significant, future studies of personal exposure to suites of VOCs should include tracer compounds to help differentiate between these three sources.

Acknowledgements

This research was supported by the US EPA grant number R82-6788-010. The PMF model was used under a licensing agreement with Pentti Paatero of the University of Helsinki. We are grateful to Dr. Ronald Henry for his assistance with UNMIX and to Dr. Pentti Paatero for his assistance with PMF.

Although the research described in this article was funded by the US EPA, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Anderson, M.J., 2001. Source apportionment of toxic volatile organic compounds. M.S. Thesis, Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO.

Anderson, M.J., Miller, S.L., Milford, J.B., 2001. Source apportionment of exposure to toxic volatile organic

compounds using positive matrix factorization. *Journal of Exposure Analysis and Environmental Epidemiology* 11, 295–307.

Clayton, C.A., Perritt, R.L., 1993. Data base development and data analysis for California Indoor Exposure Studies. California Air Resources Board, Report No. A133-187.

Daisey, J.M., Mahanama, K.R.R., Hodgson, A.T., 1994. Toxic volatile organic compounds in environmental tobacco smoke: emissions factors for modeling exposures of California populations. California Air Resources Board, Report No. A133-186.

EPA, 2000. Unified air toxics website: the pollutants. US Environmental Protection Agency, <http://www.epa.gov/ttn/uatw/> (accessed June 2000).

Henry, R.C., 2000. UNMIX Version 2 Manual. US Environmental Protection Agency, March.

Henry, R.C., Kim, B.M., 1990. Extensions of self-modeling curve resolution to mixtures of more than three components—finding the basic feasible region. *Chemometrics and Intelligent Laboratory Systems* 8, 205–216.

Henry, R.C., Lewis, C.W., Collins, J.F., 1994. Vehicle-related hydrocarbon source compositions from ambient data: the GRACE/SAFER method. *Environmental Science and Technology* 28, 823–832.

Lin, C., Milford, J.B., 1994. Decay-adjusted chemical mass balance receptor modeling for volatile organic compounds. *Atmospheric Environment* 28, 3261–3276.

McKone, T., 1987. Human exposure to volatile organic compounds in household tap water: the indoor inhalation pathway. *Environmental Science and Technology* 21, 1194–1201.

Miller, S.L., Anderson, M.J., Daly, E.P., Milford, J.B., 2002. Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data *Atmospheric Environment* 36, 3629–3641.

Mølhave, L., 1982. Indoor air pollution due to organic gases and vapours of solvents in building materials. *Environment International* 8, 117–127.

NERL, 2000. THERdbASE exposure assessment software. National Exposure Research Laboratory, US Environmental Protection Agency. <http://www.epa.gov/nerl/heasd/therd-home.htm> (accessed July 2000).

Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111–126.

Pandian, M.D., Bradford, J., Behar, J.V., 1989. THERDBASE: Total human exposure relational database. In: *Proceedings of the EPA/AWMA Specialty Conference on Total Exposure Assessment Methodology*, A New Horizon. Air and Waste Management Association, Pittsburgh, PA.

Rappaport, S.M., Selvin, S., Waters, M.A., 1987. Exposures to hydrocarbon components of gasoline in the petroleum industry. *Applied Industrial Hygiene* 2, 148–154.

Scheff, P.A., Porter, J.A., Doskey, P.V., 1991. Improvement of VOC source fingerprints for vehicles and refineries. Paper No. 91-79.5, *Proceedings from the 84th Annual Meeting and Exhibition*. Air and Waste Management Association, Vancouver, BC.

Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter

- pollution in metropolitan Boston. *Atmospheric Environment* 19, 9–25.
- Wallace, L.A., 1987. The total exposure assessment methodology (TEAM) study: summary and analysis, Vol. I. US Environmental Protection Agency, Report No. EPA/600/6-87/002a.
- Wallace, L.A., 1989. The total exposure assessment methodology (TEAM) study: an analysis of exposures, sources and risks associated with four volatile organic chemicals. *Journal of the American College of Toxicology* 8, 883–895.
- Wallace, L.A., 1991. Comparison of risk from outdoor and indoor exposure to toxic chemicals. *Environmental Health Perspectives* 95, 7–13.
- Wallace, L.A., Pellizari, E.D., Lederer, B., Zelon, Z.H., Sheldon, L., 1987. Emissions of volatile organic compounds from building materials and consumer products. *Atmospheric Environment* 21, 385–393.
- Wallace, L.A., Pellizari, E.D., Hartwell, T.D., Davis, V., Michael, L.C., Whitmore, R.W., 1989. The influence of personal activities on exposure to volatile organic compounds. *Environmental Research* 50, 37–55.
- Watson, J.G., Robinson, N.F., Chow, J.C., Henry, R.C., Kim, B.M., Pace, T.G., Meyer, E.L., Nguyen, Q., 1990a. The USEPA/DRI chemical mass balance receptor model. *Environmental Software* 5, 38–49.
- Watson, J.G., Robinson, N.F., Chow, J.C., Henry, R.C., Kim, B.M., Nguyen, Q., Meyer, E.L., Pace, T.G., 1990b. Receptor model Technical Series, Vol. III, CMB7 User's Manual. US Environmental Protection Agency, Report No. EPA-450/4-90-004.
- Yakovleva, E., Hopke, P.K., Wallace, L.A., 1999. Receptor modeling assessment of particle total exposure assessment methodology data. *Environmental Science and Technology* 33, 3645–3652.